.Species distribution modeling of the threatened Blanding's turtle's (*Emydoidea blandingii*) range edge as a tool for conservation planning.

Final Report

K.M. Stryszowska
PhD candidate
Institute for a Sustainable Environment, Clarkson University

T.A. Langen
Professor
Biology Department, Clarkson University

G. Johnson
Professor
Biology Department, The State University of New York at Potsdam

A report prepared for
St. Lawrence River Research and Education Fund and
Massena Remedial Action Committee

January 2015

Executive Summary

This report provides the results of a study to identify ecological and landscape variables responsible for creating the range edge of the threatened Blanding's turtle, *E. blandingii* in northeastern New York State.

E. blandingii, is a semiaquatic freshwater turtle of the northern United States and southeastern Canada that is of conservation concern across its range; it is currently listed as 'threatened' in New York State and 'globally endangered' by the IUCN Red List. Loss of ephemeral wetlands and decreases in permanent wetland quality and quantity, and possibly habitat fragmentation caused by roads, have contributed to the historical decrease in *E. blandingii* turtle distribution and abundance. The species appears to be near the eastern limits of its contiguous range in northern New York, although widely-distributed across the St. Lawrence River (SLR) in Ontario, Canada; disjunct populations occur in southeastern New York, Massachusetts, New Hampshire, Maine, and Nova Scotia. Because of uncertainties about its historical distribution, the extent and stability of the full range of *E. blandingii* is currently unclear

The objective of our study was to determine what environmental, ecological, and landscape variables are responsible for creating the *E. blandingii* habitat range edge evident in the SLR Valley. We also investigated whether suitable *E. blandingii* habitat exists outside the species' current range in the SLR Valley. Our data are intended to address the information needs of agencies working to restore and manage *E. blandingii* turtle habitat and populations along the SLR Valley of New York. Such information will be necessary to develop a management plan that is effective at addressing threats to the species and providing habitat management recommendations that lead to its recovery in New York.

Using the mapping and analysis software ArcGIS, we have constructed habitat suitability models at two different scales (250 m and 8,000 m) based on 211 turtle occurrence records, 113 absence records, and 11 predictor variables that are a combination of features of the landscape, turtle ecology, and environmental variables. Our results indicated that a distinct environmental range edge in the SLR Valley is associated with an elevation gradient. Models also projected the presence of suitable habitat outside of the current range, including locations with known disjunct populations.

We conclude that the range edge of *E. blandingii* in northeastern New York is limited by elevation, a natural topographic barrier. We suggest that conservation of *E. blandingii* in the SLR Valley should focus on understanding local population dynamics and managing current populations, which may be undergoing fragmentation. The information provided by the species distribution models can be used to refine the steps taken to conserve, restore, and manage *E. blandingii* habitat, both locally and regionally. Future survey sites and monitoring efforts can also be targeted using these models. We also conclude that while our models identified suitable habitat outside of the current *E. blandingii* range, one must be cautious making conservation decisions based on their forecasts. Nevertheless, our results indicate that projections can provide clues to the historical species distribution and potential for species range expansion.

Section 1: The first section of this report is a scientific manuscript (currently under review for publishing in a scientific journal) addressing not only the original objectives of our SLRREF-funded study but also an additional objective of evaluating two methods of species distribution modeling. The manuscript provides the technical details of the methods used to build the models, and offers results on the predicted distribution and habitat suitability for Blanding's Turtle in both the St. Lawrence River Valley of New York and in entire New York State.

Section 2: The second section of this report provides a focused discussion on how the results of our species distribution modeling study can be used in the St. Lawrence River Valley and the Massena Area of Concern and applied to aid local conservation efforts on behalf of Blanding's Turtle.

Table of Contents

Section 1

Species Distribution Modeling of Blanding's Turtle - Manuscript

Species distribution modeling of the threatened Blanding’s turtle’s (*Emydoidea blandingii*) range edge as a tool for conservation planning.

Kinga M. Stryszowska [a], Glenn Johnson [b], Tom A. Langen [c]

[a] Institute for a Sustainable Environment, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699, USA, stryszk@clarkson.edu
[b] Department of Biology, The State University of New York at Potsdam, 44 Pierrepont Avenue, NY 13676, USA, johnsong@potsdam.edu
[c] Department of Biology, Clarkson University, 8 Clarkson Avenue, Potsdam, NY 13699-5805, USA, tlangen@clarkson.edu, phone: 315-268-7933

Abstract

The delineation of a species range is challenging because of the number of interacting factors at multiple spatial scales affecting a species’ distribution. Species distribution models (SDM) can be used to identify factors that are most associated with a species’ presence and thus potentially define a range edge. We evaluated the predictive success of two popular SDM approaches, maximum entropy models (Maxent) and generalized linear models (GLM), at determining the range edge for the threatened Blanding’s turtle, *Emydoidea blandingii*, in northeastern New York. This is the first study comparing performance of two SDM approaches using a large sample size of presence/absence records and presence/background records in a small geographic area (585,000 ha) at the range edge of a rare species. Using the mapping and analysis software ArcGIS, we constructed and validated SDMs using presence/absence records (GLM) and presence/background records (Maxent) with 11 environmental predictor variables. We found that Maxent was more successful at predicting habitat suitability than GLM. Our results also indicated that a distinct environmental range edge is associated with an elevation gradient. Both GLM and Maxent models also projected the presence of suitable habitat outside of the current range, including locations with known disjunct populations. We conclude that a presence/background SDM approach like Maxent is valid when accurate data on locational absences are lacking, as is typical for rare, cryptic species. SDM used to understand the factors shaping the range edge can aid at planning habitat conservation and management of threatened species.

Keywords

Emydoidea blandingii
Species Distribution Model
Freshwater turtle
Range edge
Species conservation

Introduction

Species’ distributions, on both a global and local scale are limited in part by the demographic responses of populations to spatial variation in environmental factors (Gaston 2009; Geber 2008; Sexton et al. 2009). As environmental factors become unfavorable, species respond by a decrease in population density until a range edge is created. Spatial environmental variation, however, is only a single element among the complex, dynamic factors that determine a species' range. Evolutionary history, population dynamics, species vagility, physical barriers, community interactions, and resource availability interact at multiple spatial scales to create a range edge (Baselga et al. 2012; Brown et al. 1996; Holt and Keitt 2005). Range edge dynamics of rare species are particularly challenging to study because of sparse reliable data; rare species typically have low detection probabilities, so determining presence versus absence at localities is very difficult (Engler et al. 2004; Manel et al. 2001; Marini et al. 2010). Nevertheless, rare species have a greater need for range edge delineation than common species because knowing the physical and environmental limits to population persistence is requisite for successful conservation planning.

Species distribution models (SDMs) are useful tools for delineating species ranges and identifying environmental factors affecting distribution (Arntzen and Espregueira Themudo 2008, Seabrook et al. 2014; Tarkhnishvili et al. 2009). SDMs extrapolate species habitat tolerances and preferences to predict the geographical distribution of species in areas that have not been surveyed (Franklin 2009; Guisan and Zimmermann 2000). SDM algorithms can identify factors most associated with species occurrence, including factors relevant to conservation planning. Whereas SDMs have successfully modeled invasive species distributions (Crall et al. 2013; Thuiller et al. 2005) and species range responses to climate change (Austin and Van Niel 2011; Eskildsen et al. 2013; Seabrook et al. 2014), few studies have explicitly modeled population dynamics at a range edge (Eskildsen et al. 2013; Tarkhnishvili et al. 2009; Williams-Tripp et al. 2012).

The predicted species distributions produced by SDMs typically have a higher resolution than global species range maps, but are seldom used to provide predictions at fine scales (e.g. less than 100 ha) and for small geographic regions, even for small-bodied organisms for which such resolution would be ideal (Chefaoui and Lobo 2008; Marini et al. 2010; Pearson et al. 2007). This is likely the consequence of the paucity of very high resolution (30 m or less) GIS coverage maps and inherent unreliability of building models from variables that lack spatial variability at fine scales and small geographic extents (Khatchikian et al. 2011; Wang et. al. 2012).

Species occurrence locality data used to build SDMs are typically presence-only data since true absences are difficult to determine, especially for low-density and cryptic species (Elith et al. 2006; Segurado and Araujo 2004). To achieve near-certainty of absence requires repeated, long term surveys of putative absence sites, and for many studies such intensive surveys are not feasible (MacKenzie et al. 2002). Some studies have found that using absence records produces more accurate models when compared to presence-only records (Brotons et al. 2004), while others have found that absence records do not increase precision (Wintle et al. 2005) or even produced less accurate models (Rupprecht et al. 2011).

Freshwater turtles are undergoing steep population declines worldwide (Gibbons et al. 2000; Millennium Ecosystem Assessment, 2005). The Blanding's turtle, *Emydoidea blandingii*, is a semiaquatic freshwater turtle of the northern United States and southeastern Canada that is of conservation concern across its range; it is currently listed as 'threatened' in New York State (Ross and Johnson 2013) and 'globally endangered' by the IUCN Red List. The species is notable for long overland movements among permanent and ephemeral wetlands within a season (Congdon et al. 2008; Millar and Blouin-Demers 2011; Refsnider and Linck 2012). Loss of ephemeral wetlands and decreases in permanent wetland quality and quantity, and possibly habitat fragmentation caused by roads, have contributed to the historical decrease in *E. blandingii* turtle distribution and abundance (Congdon et al. 2008; Ernst and Lovich 2009; Compton 2007; Johnson and Crockett 2009).

Because of uncertainties about its historical distribution, the extent and stability of the full range of *E. blandingii* is currently unclear. The first records of *E. blandingii* in northeastern New York State, in the headwaters of the St. Lawrence River (SLR), were reported in 1951 (Werner 1959). Since that time occurrence records have increased in number and spatial extent in the region as a result of increased research and volunteer survey effort (Johnson 2012; Ross and Johnson 2013). The species appears to be near the eastern limits of its contiguous range in northern New York, although widely-distributed across the SLR in Ontario, Canada, and disjunct populations occur in southeastern New York, Massachusetts, New Hampshire, Maine, and Nova Scotia. While there is evidence that the edge populations are increasingly isolated genetically from the core Midwestern U.S. populations (Mockford et al 2007; McCluskey et al. 2015), whether the turtle's range is presently expanding or retracting is unknown.

Studies that model the habitat associations of *E. blandingii* and responses to landscape fragmentation within the core range conclude that this species most frequently occurs in large lakes and marshes (Rizkalla and Swihart 2006; Attum et al. 2008) distant from roads (Attum et al. 2008), within landscapes with heavy forest cover (Quesnelle et al. 2013; Millar and Blouin-Demers 2012). Additionally, in Ontario Millar and Blouin-Demers (2012) found that habitat suitability increased with average air temperature (up to 25º C). In aggregate, these studies indicate that *E. blandingii* may not only be vulnerable to reductions in habitat size and increases in habitat fragmentation, but also respond to changes in climate.

The importance of spatial scale cannot be overemphasized when delineating a range edge. Organisms respond to their environment at a variety of scales and it is difficult to determine *a priori*, which scale(s) may be most appropriate for a species. When building an SDM, most modelers use scales that reflect vagility of individuals or their propagules (Attum et al. 2008; Mateo-Tomas and Olea 2009; Quesnelle et al. 2013). For example, Millar and Blouin-Demers (2012) used three ecologically meaningful scales to model the distribution of *E. blandingii* in Ontario, Canada, with the largest scale (1,000 m) analogous to home range length. Quesnelle et al. (2013) similarly used a scale of 1.2 km. We posit that a scale larger than home-range size of individual turtles may be informative when trying to understand the influence of landscape and climate on species distribution. A macro scale is more appropriate when modeling climatic responses because climatic variables rarely exhibit extreme variation at fine scales (Austin and Van Niel 2002; Pearson et al. 2004). A larger scale can also represent distal variables affecting turtle distribution including biotic interactions at the community or landscape scale and metapopulation dynamics (Austin 2002, Marchand and Litvaitis 2004, Heikkinen et al. 2007).

Our study had two objectives: (1) to evaluate the effectiveness of two popular SDM methods at predicting occurrences of a rare species at the edge of its range when using presence-only versus presence-absence records, and (2) to use SDM to determine which environmental and landscape factors affect the distribution of *E. blandingii*, especially those contributing to its range limit in northeastern New York. We used spatially-extensive survey data on *E. blandingii* distribution in northeastern New York, regional climate data, and high-resolution land cover data to develop SDMs using maximum entropy (Maxent) and generalized linear models (GLM). We hypothesized that *E. blandingii* populations in the SLR Valley of northeastern New York are constrained by the abrupt elevation rise at the Adirondack Mountains along the southern border of the valley and by the prevalence of habitat fragmentation from agricultural and residential development, and roads within the valley. We also hypothesized that an SDM algorithm that incorporates true absence records (i.e. GLM) would perform better at predicting species occurrences at a range edge than one that doesn't (i.e. Maxent). Understanding the environmental and climatic factors determining the range edge of *E. blandingii* can shed light on the stability of this species' current range and guide conservation and restoration planning throughout it. Understanding how well different SDM approaches perform at range edges of rare species is essential if SDM is to be used as a tool for conservation planning.

Methods

Study area

The study area is the 650,000 ha SLR Valley along the northeastern border of New York State, and bordered by the Adirondack Mountains to the southeast. It is characterized by a temperate climate with little variation across its span. Based on the 2011 National Land Cover Data (NLCD) the region is dominated by a mix of agricultural uses and northern hardwood-conifer forest; emergent and shrub wetlands comprise 3.5% of the landscape (Jin et al. 2013). The 585,000 ha region selected for SDM was delineated by taking the longest distance between two adjacent *E. blandingii* occurrence records (21.1 km) and buffering all occurrence records by that distance (after Aitken et al. 2007). The northern edge was defined by the New York State border; *E. blandingii* distribution in adjacent Ontario, Canada was ignored due to incommensurability of landscape, environmental, and survey data (see Millar and Blouin-Demers 2012).

Data sources

Our *E. blandingii* occurrence records were obtained from a fourteen year (1999 - 2013) regional survey, performed by setting turtle hoop nets in wetlands for multiple nights; effort ranged from 15 to 2,170 trap-nights per site (see Johnson and Conrad 2012). Survey sites were selected to provide thorough geographic coverage, and surveys targeted accessible woody (forested or shrub) wetlands, a habitat preferred by *E. blandingii*. Wetlands near ad hoc turtle detections (e.g. road crossings or road-kill records, reports from the public) were also surveyed. Among the 228 surveyed sites, 87 sites had one or more *E. blandingii* detection. Trapping records were supplemented with road crossing records and individuals collected by hand during surveys, which resulted in an additional 102 occurrence records. To reduce spatial autocorrelation, only occurrences that were at least 100 m apart were included in the final analysis. For the coarser spatial scale SDMs, only one occurrence record was retained in each 800 m x 800 m raster cell. Out of the final 211 occurrence records, 99 were road crossings, 66 were hand collections, and 46 were trap records. We constructed models by randomly selecting 75% of the occurrence records and leaving 25% for independent validation. To calculate the probability of a true absence if there was no detection during a survey, the proportional probability of detection per trap-night (P_{site}) was first estimated for each of eight sites previously known to have *E. blandingii* present and for which survey effort was at least 300 trap-nights (Kery 2002). The number of turtles captured at each site (N) was divided by the total trap nights (T) of that site: $P_{site} = N / T$. The proportional probabilities for the eight sites were then averaged (P_{mean} = 0.0169 ± SD 0.0110). The proportional probability of a true absence (q) was then calculated: $q = 1 - (P_{mean}^{0} * (1 - P_{mean}^{T}))$. To retain a suitable number of sampled sites for SDM, the proportional probability of a true absence was cut off at 0.3 (70% chance of occurrence despite no detections) which resulted in retaining 113 out of 131 no-detection sites.

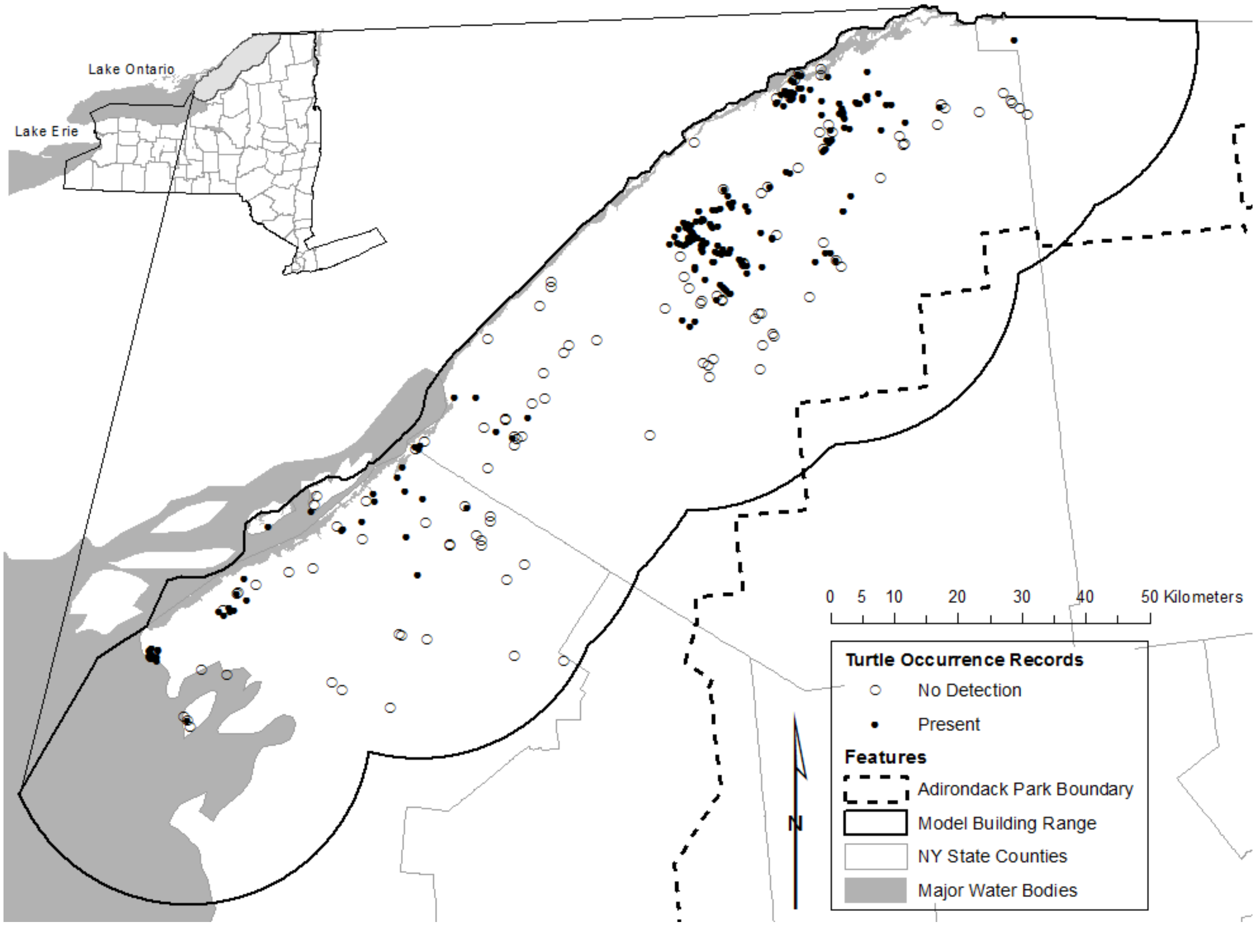


Figure 1: Regional limits for SDM in the St. Lawrence River Valley of New York, also indicating *E. blandingii* survey locations.

We used 11 environmental putative predictor variables associated with three principal factors: climate, land cover, and topography (Table 1). We verified that variance inflation factors were low (<5) to reduce multicollinearity among them. Final predictor variables were chosen based on our knowledge of *E. blandingii* ecology and comparable turtle models (Rizkalla and Swihart 2006; Quesnelle et al. 2013; Millar and Blouin-Demers 2012; Attum et al. 2008). A forest canopy density (deciduous and coniferous) variable was derived from the 2011 NLCD (Jin et al. 2013), and a hardwood forest cover variable from the 2008 Northeastern Terrestrial Wildlife Habitat Classification map (Gawler 2008) by extracting the dominant forest habitat type (Appalachian (Hemlock)-Northern Hardwood Forest). We derived two wetland variables (emergent wetlands and forested/shrub wetlands) from the National Wetland Inventory (NWI;

USFS 1983), and an open water variable from a combination of NLCD and NWI. We derived a corn/alfalfa land cover variable from the 2010 USDA cropland dataset (USDA National Agricultural Statistics Service Cropland Data Layer. 2010). Road density measures were calculated using the summed length of road segments from the 2013 TIGER roads polyline file (U.S Census Bureau 2013); stream density was calculated using the Northeast Aquatic Habitat stream polyline file (Olivero and Anderson 2008). We used the 2009 USGS National Elevation Dataset at 1/9 arc second (30 m) resolution for the elevation variable (Gesch 2002, Gesch 2007). We calculated mean monthly precipitation (mm) and mean maximum monthly temperature (°C) using the WorldClim 1960-1990 dataset at 800 m resolution (Hijmans et al. 2005); monthly rasters were averaged across the *E. blandingii* active period (April-October). Because of the coarse resolution of the two climate variables, they were not used at the 250 m scale of modeling.

To evaluate the effects of different landscape scales on the occurrence of *E. blandingii*, we extracted environmental data within two circular buffers around each presence/absence point: 250 m (raster resolution of 30 m) and 8,000 m (rasters re-sampled using bilinear interpolation to 800 m). Environmental variables were read as either the sum of pixels or mean of pixel values within a buffer area. The 250 m radius corresponded to a previously used buffer length intended to match *E. blandingii* mean daily movement distances, and thus turtles' direct interactions with the landscape (Millar and Blouin-Demers 2012). The 8,000 m buffer was intended to capture indirect landscape-scale factors. For both GLM and Maxent, and at each buffer scale, we ran three replicate models using different, random combinations (75% train/25% test) of occurrence records. All spatial analyses were conducted using ArcGIS Desktop 10.2.1 (ESRI 2013) using the North American Datum 1983 and NAD 1983 UTM Zone 18 N projection.

Table 1: Eleven predictor variables used to build the SDMs for *E. blandingii* in New York.

Variables (units)	250 m scale		8 000 m scale	
	Range	Median	Range	Median
Mean monthly precipitation (April-October, mm)	NA	NA	76.7 - 88.6	82.1
Mean monthly maximal temperature (April-October, ºC)	NA	NA	19.7 - 20.3	20.0
Elevation (m)	60.6 - 191.9	94.9	59.0 - 224.5	99.8
Road density (km/km^2)	0 - 8.4	2.5	0.5 - 3.2	1.0
Stream density (km/km^2)	0 - 4.9	0	0.1 - 1.1	0.7
Land cover (%)				
Forested/shrub wetland	0 - 84.4	12.4	0.3 - 25.5	13.0
Emergent wetland	0 - 88.1	0	0.4 - 4.0	1.3
Canopy density	0 - 82.1	20.6	3.8 - 71.7	30.3
Hardwood forest	0 - 77.5	10.6	1.4 - 53.1	17.8
Open water	0 - 61.5	0	0.5 - 75.5	3.8
Alfalfa/corn	0 - 50.5	0.5	0.2 - 20.0	3.5

Model building

We compare the two most popular methods of SDM: GLM and Maxent. Both methods have been found to be accurate at predicting species occurrences, but differ importantly in the type of data they use (Elith et al. 2006; Pearson et al. 2007; Khatchikian et al. 2011). GLM is an extension of linear regression that can model binomial data distributions (Guisan et al. 2002); for this reason, GLMs are used when both presence and absence records are available for comparison (Franklin 2009; Rupprecht et al. 2011). Maxent is a machine learning model that uses presence records compared to a random sample of background locations to find the probability distribution of maximum entropy (i.e., closest to uniform) without over-fitting the model (Phillips et al. 2006).

GLM

All GLMs were run in R version 3.0.3 (R Core Team 2014) using the function *glm* with a binomial distribution and a logit link function. For both scales, we created all possible model subsets and selected the model with the lowest Akaike Information Criterion (AIC). Variable contributions were determined by standardizing all predictor variables (z-scores); the highest slope coefficients were judged the most influential. The log odds ratio was converted into probability of occurrence (y) from 0 (low) to 1 (high) and imported into ArcMap using the equation: $y = 1 / (1 + \exp - (a + \Sigma\, x{*}b))$, where ($a$) is the intercept, ($x$) is the regression coefficient for each model variable and (b) is the variable raster. The final model was validated using 25% (53 presence and 28 absence) of the original presence/absence records.

Maxent

Maxent version 3.3.3 was used to build an SDM on 159 presence records. We used most of the default parameters, with a few modifications (Phillips et al. 2006; Phillips and Dudik 2008). Maxent was set to uniformly sample 10,000 background locations across the study region, intended to characterize the distribution of environmental parameter values. Because trapping surveys were not a random sample of localities within the region, but rather targeted woody wetlands and tended to be near roads, we implemented a wetland-road bias to select background points for the model. The bias file was created by buffering all roads and all forested and shrub wetlands by the mean distance of an occurrence record to these features. We increased the number of iterations over the Maxent default to 5,000 to allow the model adequate time for convergence. We also adjusted the default prevalence value of 0.5 to 0.382 to better represent the prevalence of this rare species (Merow et al. 2013; Elith et al. 2011). Prevalence equals the number of surveyed sites with detections (87) divided by total surveyed sites (228). Variable relative contributions to the SDM were inferred by the increase in model gain when added. Contribution was also judged by inspecting ‘jackknife’ contribution plots. Model predictions were imported into ArcMap in the logistic format, providing a predicted spatial probability of occurrence from 0 (low) to 1 (high). As with GLM, 25% of the presence records were used to validate the model, and the same validation points were used as in the GLM model validation.

Model evaluation

We used area-under-the-curve (AUC) of the receiver-operating-characteristic (ROC) curve as the primary method of model evaluation. The advantage of AUC is that it does not require specification of a threshold to convert continuous probability data into a binary output. The AUC test is derived from the ROC curve, which is plotted using sensitivity (proportion of presences correctly predicted) and specificity (proportion of absences correctly predicted) (Pearson 2007). AUC ranges from 0.5 (models no better than random) to 1 (perfect discrimination). An accepted rule-of-thumb is models with AUC values above 0.75 are considered informative (Eskildsen et al. 2013; Swets 1988). AUC was used to evaluate the fit of the final models to the building points and to evaluate the models' successes in predicting the validation points. For the purpose of additional model evaluation and interpretation via habitat suitability maps, thresholds were set for both GLM and Maxent models to convert continuous probability data into a binary format. For both models we chose a threshold that maximized the sum of sensitivity and specificity (Liu et al. 2005; Jimenez-Valverde and Lobo 2007). Replicate binary map rasters were averaged for a single display map.

Projection

Our SDM results were used not only to predict the distribution of *E. blandingii* within our study region, but also projected to the rest of New York State to evaluate whether suitable habitat for *E. blandingii*, as indicated by our models, existed outside of the modeled range. We compared these model projections to the distribution of known *E. blandingii* populations throughout the state.

Results

Both SDM methods performed reasonably well at both scales, according to our acceptance criterion of AUC = 0.75, and closely fitted the training points (Table 2). The mean training AUC value among models was 0.959 ± SD 0.004 (Maxent) and 0.855 ± 0.018 (GLM). The mean validation AUC value was 0.911 ± 0.027 (Maxent) and 0.661 ± 0.159 (GLM). GLM training and validation AUC values were highest at the 8,000 m scale, whereas for Maxent, training AUC values were the same at both scales and validation AUC was highest at the 250 m scale (Table 2). Overall, Maxent models, using presence locations with background samples and a bias file, performed better both in model fit and validation than the presence/absence-based GLM models (Table 2).

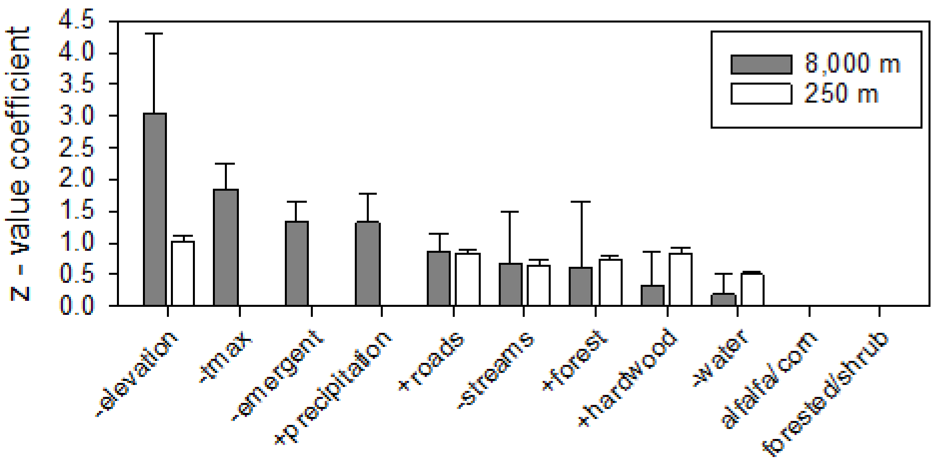
Table 2: Performance of GLM and Maxent models at 250 m and 8,000 m scales. The mean ± SD AUC is reported for model fit to 75% training data and to 25% independent validation data (N=3).

Model	Scale	Mean AUC	
		Training	Validation
GLM	250 m	0.846 ± 0.008	0.612 ± 0.215
	8000 m	0.864 ± 0.022	0.710 ± 0.098
Maxent	250 m	0.959 ± 0.004	0.913 ± 0.020
	8000 m	0.959 ± 0.005	0.909 ± 0.036

Variable contributions

When using GLM, the most important variable averaged across both scales was elevation, followed by mean monthly maximum temperature and mean monthly precipitation (Figure 2); the latter two were only used at the 8,000 m scale because of their course resolution. Habitat suitability for *E. blandingii* was lower with higher elevations and mean temperatures, and higher with higher precipitation. Other important explanatory variables were road density, forest canopy cover and stream density. Forested/shrub wetland cover and alfalfa/corn cover were not included in any of the GLM models at any scale (Figure 2).

When using Maxent, the most important explanatory variable averaged across both scales was elevation, followed by road density; the latter was a very important predictor at the 250 m scale and was trivial at the 8,000 m scale. As with GLM, *E. blandingii* habitat suitability was lower at higher elevations and, at the 250 m scale, decreased with increasing road density (Figure 2). Other important explanatory variables were forested/shrub wetland cover and stream density. When building models, Maxent uses all provided variables, excluding none from the final model. The consistently least important variables were forest canopy and alfalfa/corn cover. The effect of some variables on habitat suitability was difficult to interpret from the jackknife contribution plots.


(A) GLM

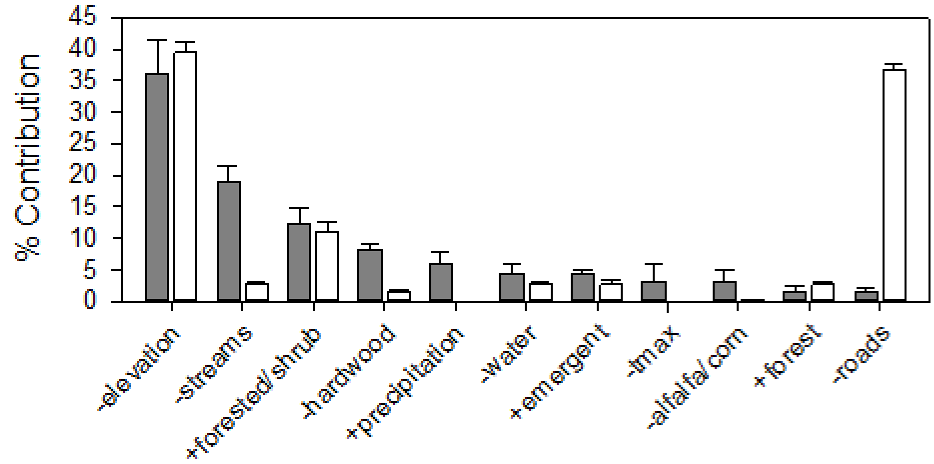

(B) Maxent

Figure 2. Mean importance of predictor variables for (A) GLM and (B) Maxent models at 250 m and 8,000 m scales. Error bars represent one standard deviation. Symbols (+/-) in front of variables represent direction of influence on habitat suitability. Missing bars represent variables not included in the models. Importance in GLM models is expressed in terms of the regression coefficient value after z-value standardization; higher coefficients are more important. Importance in Maxent models is expressed as percent contribution to the model.

The most marked difference between Maxent and GLM models was that temperature was of very little importance in the Maxent models, whereas it was the second highest contributing variable in the GLM models (Figure 2). Hardwood forest cover was of much greater importance in GLM models, whereas forested/shrub wetland cover was of greater importance in Maxent models. While road density was important at the 250 m scale for both Maxent and GLM, the effect it had on habitat suitability differed (Figure 2). In the Maxent models, the direction of the effect of some variables on habitat suitability changed with changing scales, whereas with GLM, the effect was consistent across scales.

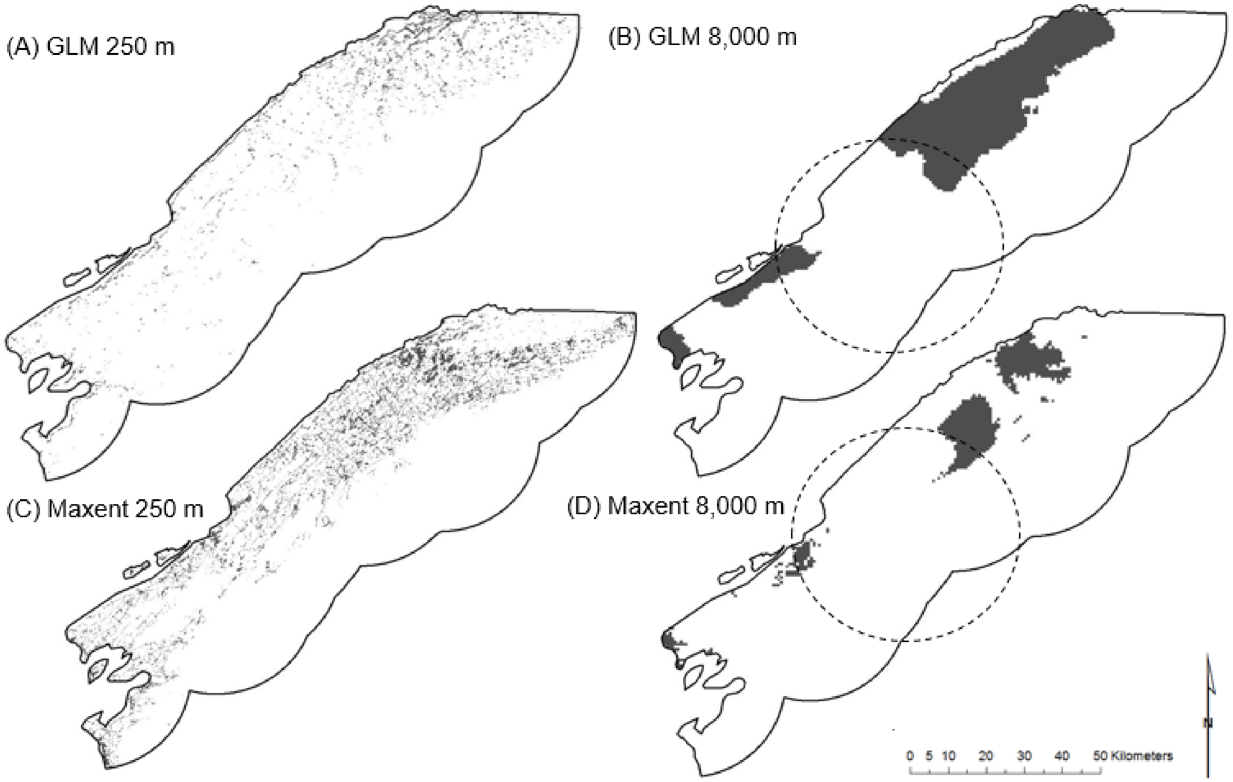


Figure 3: Probability of occurrence of *E. blandingii* across the building extent in northern New York using (A) GLM with a 250 m buffer and (B) 8,000 m buffer, and (C) Maxent with a 250 m buffer and (D) 8,000 m buffer. Gray areas indicate high probability of occurrence. Circled area is a predicted gap between two areas of high probability of occurrence.

Habitat suitability predictions within the study region

At the 250 m scale, both algorithms predicted small patches of high occurrence probability across the study area. Denser clusters were evident in the northeast section of the study region (Figure 3). At the 8,000 m scale both algorithms predicted high probability of occurrence in the same locations and an area of low probability of occurrence in the middle of the region (Figure 3). The most notable difference between Maxent and GLM model predictions was evident at the 8,000 m scale where Maxent predicted small patches of high probability of occurrence while GLM predicted large areas of high probability of occurrence, especially in the northeast of the region. At all scales for both Maxent and GLM predictions, probability of *E.*

blandingii occurrence decreased when moving in the southeastern direction away from the SLR Valley and toward the Adirondack Mountains.

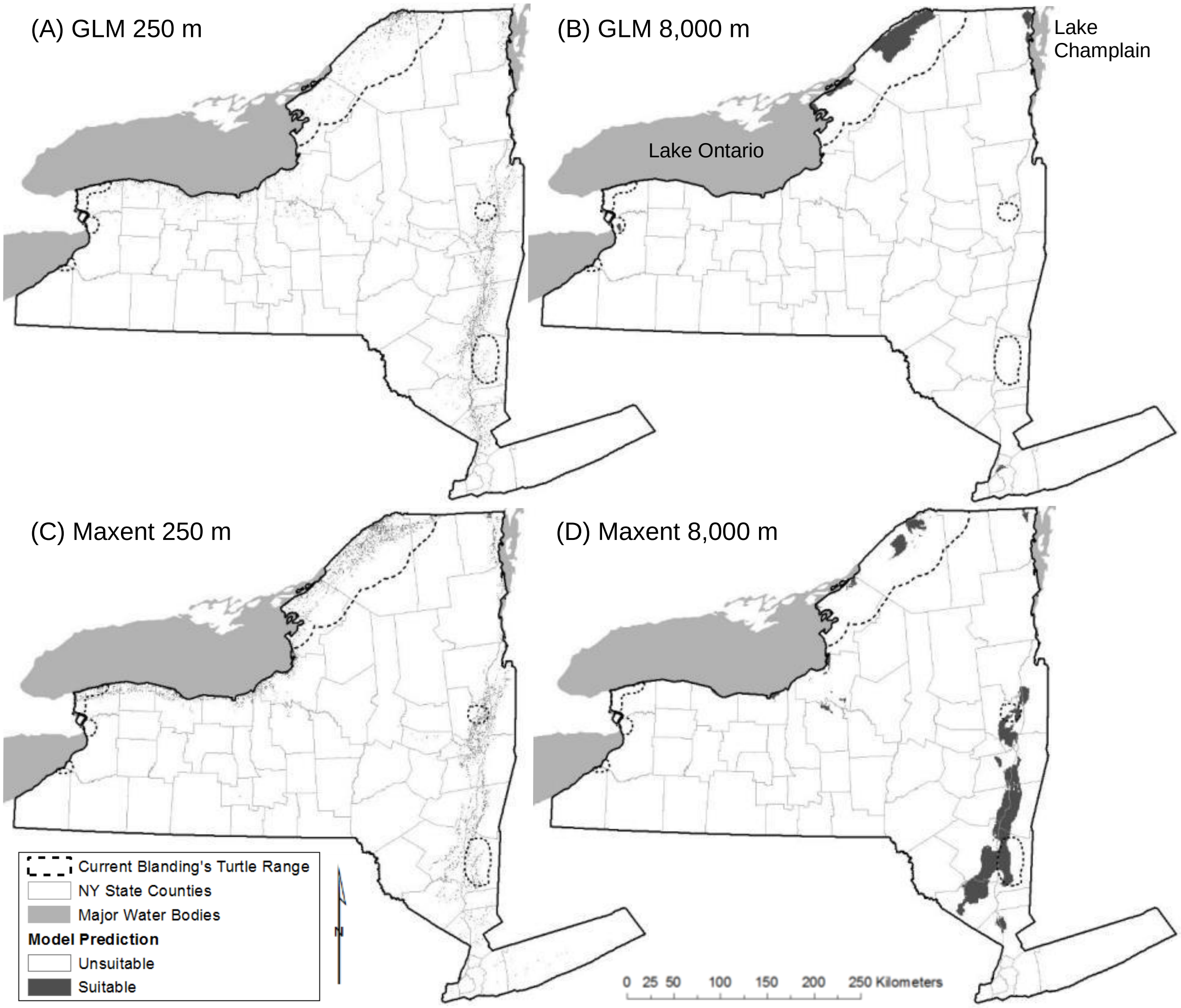


Figure 4: Projected areas of high habitat suitability for *E. blandingii* outside of the model building extent using (A) GLM with a 250 m buffer and (B) 8,000 m buffer, and (C) Maxent with a 250 m buffer, and (D) 8,000 m buffer. The dotted line delineates the current known distribution of *E. blandingii* in New York State.

Habitat suitability predictions projected outside the study region

When projecting outside of the range, GLM at the 250 m scale performed differently than at the 8,000 m scale. At the 250 m scale, high habitat suitability was predicted along Lake Champlain and down the Hudson River Valley. This predicted distribution encompassed two known, disjunct populations of *E. blandingii* (Ross and Johnson 2013). High suitability was also predicted along the southern shore of Lake Ontario to the western border of NY State, encompassing poorly surveyed disjunct populations in that region (Figure 4A). At the 8,000 m scale, high habitat suitability was predicted for few sites outside the current range, no longer classifying the region of the Hudson River Valley populations as suitable. The 8,000 m scale model predicted small areas of high suitability along the western shore of Lake Champlain and along the eastern shore of Lake Ontario, where there are no records (Figure 4B).

At the 250 m scale, the Maxent prediction was very similar to the GLM prediction, identifying suitable habitat along the southern shore of Lake Ontario and the Hudson River Valley. At the 8,000 m scale, similar to GLM, high habitat suitability areas were predicted for the western shore of Lake Champlain and eastern shoreline of Lake Ontario. Unlike GLM, large patches of high suitability habitat were predicted for almost the entire Hudson River Valley, encompassing the two regions with documented extant populations (Figure 4D).

Discussion

This is the first study comparing two SDM algorithms using a large sample size of presence/absence records and presence/background records in a small geographic area at the range edge of a rare species. Statistical evaluation of model fit and independent validation, in combination with visual assessment of predictive maps, indicates that the presence-only, machine learning method Maxent is better at characterizing habitat suitability for *E. blandingii* both within the modeled region and projected outside of the region. Our results also support our prior observations that the range of the *E. blandingii* in northeastern New York State is limited primarily by elevation.

Maxent vs GLM

Our Maxent models had consistently higher AUC values than GLM at both spatial scales and both for model fit and model validation; at least two other studies have had similar results (Khatchikian et al. 2011; Rupprecht et al. 2011). We had hypothesized that in a small geographic area at the range edge, where the gradients of environmental variables may be limited, the inclusion of absences may provide an important level of discrimination. Our results did not support our a priori expectation that a presence/absence model would have better predictive power. In general, studies comparing SDM with and without true absences have had mixed conclusions (Brotons et al. 2004; Rupprecht et al. 2011; Wintle et al. 2005).

Small study regions pose the challenge of incompletely representing the entire breadth of a species' environmental niche. Not sampling the entire gradient of the tolerances of a species can seriously bias model predictions (Hortal et al. 2008; Jimenez-Valverde et al. 2009). Populations of species of conservation concern are often located in very small geographic regions that only contain truncated ranges of environmental gradients. Populations may be disjunct or at the range edge where conditions can be very different from the range core. Our results indicate that when the environmental gradient is truncated within a study region, Maxent was the better-performing algorithm. Maxent models made more conservative, but more accurate predictions at the range edge. Moreover, when projecting beyond the modeling extent, Maxent models accurately predicted known disjunct *E. blandingii* populations in New York. Maxent, by default, selects 10 000 background points to characterize the entire study region, while in our study GLM was limited to the 113 absences we provided. These 113 locations were not a random sample of the region, but prospectively selected because they had habitat indicators associated with suitable *E. blandingii* habitat. GLM may have been over-constrained by the very low cutoff value applied to the selection of our absence records (30% probability of absence). This cutoff value may have been too low and may have decreased the precision of the GLM models by unreliably classifying unsuitable habitat (Gu and Swihart 2004). High confidence of true absence can require impractically high sample effort for a cryptic, rare species such as *E. blandingii*; for species that are more detectable, GLM with presence/absence records may perform better.

Variable contributions

We found that elevation was the most important overall predictor of *E. blandingii* habitat suitability both GLM and Maxent models and at both scales (Figures 3 and 4). Elevation is not a variable commonly used in modeling the habitat preferences of turtles (DeCatanzaro and Chow-Fraser 2010; Millar and Blouin-Demers 2012; Quesnelle et al. 2013), but in our case it was important because of the proximity of the Adirondack Mountains. There is a pronounced elevation gradient increasing from the shore of the SLR southeast into the Adirondack Mountains, and *E. blandingii* is restricted to lower elevations. The second most important variable in the GLM models was mean monthly maximum temperature, with probability of occurrence decreasing with increasing temperature (Figure 2). The apparent response of *E. blandingii* to this climate variable contrasts with Millar and Blouin-Demers (2012) in adjacent Ontario, Canada, and is also counterintuitive in relation to the response to elevation. Though *E. blandingii* thrives in the warmer Midwestern US region of its range (Congdon et al. 2008) and is likely cold-stressed along the northern range limit (Millar and Blouin-Demers, 2012), this turtle is associated with cooler areas of the SLR Valley. The response of *E. blandingii* to both elevation and temperature may well be related to a suite of factors that are associated with these two variables, e.g. soils, hydrology, land cover and use, food availability, slope of the landscape, and microhabitat characteristics (Guisan and Zimmermann 2000).

Habitat suitability predictions within the study region

Both Maxent and GLM algorithms, at both scales, identified a pronounced range edge for *E. blandingii* in northern New York that is primarily associated with elevation (Figure 2). Habitat suitability predictions indicate that populations are concentrated near the SLR, and suitable habitat is not uniform across the valley; the models identified more than half of it as unsuitable for *E. blandingii* (Figure 3). Patchy distribution is a common phenomenon at a range edge (Brown et al. 1996) and our models support this. SDM predictions corresponded closely with *E. blandingii* distribution records: for both Maxent and GLM SDMs, most turtle records are in areas of predicted high habitat suitability and there are very few records in areas predicted as low suitability. *E. blandingii* in the SLR Valley seem to be occupying all suitable habitat. These results, while providing distribution information, say nothing about population dynamics. Additional demographic information would help clarify whether occupied patches are population sources or sinks and explain micro-patterns (Brown et al. 1996; Gaston 2009).

Projection outside the study region

Model projection is a method of using SDMs to extrapolate suitability predictions into geographic areas or time periods not included in the original model construction. In general, projection is discouraged because (1) model algorithms may continue a fitted trend beyond the range of parameter values (Elith and Graham 2009) and (2) the model is trained under a combination of variables that may not be ecologically relevant to the species in distant portions of its range (Guisan and Zimmermann 2000). Projections based on models created at the range edge and in a small geographic region may be especially uncertain because variable gradients may be limited, and the species may not be at an equilibrial distribution (Eskildsen et al. 2013; Williams-Tripp et al. 2012). Nevertheless, we chose to project our SDMs to the rest of New York State to evaluate their performance in relation to several disjunct known populations of *E. blandingii* in the state, and to evaluate whether there may be suitable habitat elsewhere that could be of conservation value for this species. Many applications of SDM as a conservation tool require projection outside of a modeled region, so there is a need to evaluate the performance of SDM projection using species for which distribution in a projected area is known.

At the 250 m scale both GLM and Maxent models predicted regions of high habitat suitability that encompassed known small, disjunct populations within New York, but also included some extensive regions where there are no records of the species (Figure 4). Suitable but unoccupied habitat outside of the current range may indicate that (1) model projections are flawed, (2) undetected populations of turtles exist in these areas, (3) the species has not yet reached these suitable areas but may do so as a result of range expansion, or (4) populations have been extirpated from these areas due to historical habitat alteration, fragmentation, and degradation. We hypothesize that the regions of predicted extensive suitable habitat, the Hudson River and Lake Champlain Valleys and the Great Lakes lowlands, were once important components of the eastern *E. blandingii* range, but the historical extensive land cover and hydrological modification of the region for agriculture and industry resulted in region-wide extirpation, leaving as residuals the current small, disjunct populations.

The projections of the GLM models at the 8,000 m scale were restricted to northern New York State because of the limits imposed by the inclusion of the temperature variable. Maximum mean temperature in our training region spanned less than 1 ºC, from 19.7 ºC to 20.3 ºC (Table 1). Since highest suitability was found at the lowest temperatures, GLM models predicted areas at higher temperatures to be unsuitable. This response to temperature is likely very local and is opposite that found by Millar and Blouin-Demers (2012) in Ontario. For this reason, including temperature in projections outside of a modeled range may be problematic (Randin et al. 2006). At the 8,000 m scale Maxent predicted suitable habitat in the same locations as GLM models (Figure 4). The agreement between the two SDM algorithms is encouraging evidence that the projections are robust. Unlike GLM, Maxent projections at the 8,000 m scale were not limited by the temperature variable because temperature did not strongly contribute to the final model. At both spatial scales, Maxent predicted high habitat suitability at locations of known populations, and also in some other regions for which there are no occurrence records but it is plausible that *E. blandingii* once existed. The fact that our SDM projections make ecological sense and encompass known populations of *E. blandingii* indicate that model projection can be used as a conservation tool to locate promising sites for population surveys or suitable habitat for population restoration.

Conclusions and Management Implications

Our results indicate that the range edge of *E. blandingii* in northeastern New York is limited by elevation, so efforts to conserve *E. blandingii* should focus on understanding local population dynamics and managing habitat of populations within the current SLR Valley; conservation efforts to extend the range edge boundary via habitat management or population translocations are likely to be ineffective because of the limits imposed by the natural topographic barrier. Our model projections do indicate that potentially suitable habitat may exist in corridor-like patches outside of the current range, which suggests that the species may have once occupied a much larger region of New York State but has suffered range collapse due to habitat loss. Because SDM projections are inherently uncertain, one must be cautious making conservation decisions based on their forecasts. Nevertheless, our results indicate that projections can provide clues to the historical species distribution and potential for species range expansion, and the environmental factors that currently limit the distribution. Areas identified as suitable by projections can be targeted for future surveys and even evaluated as candidates for habitat management and population translocations to connect disjunct populations.

The best SDM method remains controversial. Our results indicate that while both Maxent and GLM are very good at predicting habitat suitability and range limits of a rare species, Maxent is better suited at making predictions using variables with weak gradients and in very small geographic areas. Maxent has been consistently shown to be a robust algorithm (Elith et al. 2006; Pearson et al. 2007; Phillips et al. 2006), but it has rarely been compared to presence/absence models (Khatchikian et al. 2011; Rupprecht et al. 2011). Our results suggest that the background selection method used in Maxent models is effective enough to replace true-absence data. Since the determination of absences requires a much more intensive sampling strategy, especially for rare and cryptic species, eliminating the need for absence data from SDMs can greatly increase the efficiency of building occurrence record databases (i.e. surveying more sites with less intensity per site), ultimately resulting in better presence-only models.

Acknowledgements

This study was supported by the St. Lawrence River Research and Education Fund (SLRREF). We are indebted to the many people who provided field assistance for this project: T. Crockett, A. Breisch, J. Ozard, A. Ross and E. McCluskey and many undergraduate students from SUNY Potsdam. We thank many private landowners for wetland access, and the New York Department of Conservation and SUNY Potsdam Research Foundation for logistical and financial support. This research was conducted in accordance with Institutional Animal Care and Use Committee protocol numbers 08-S-012, 10-F-017, and 11-S-019.

References

Aitken M, Roberts DW, Shultz LM. 2007. Modeling distributions of rare plants in the great basin, western North America. Western North American Naturalist 67:26-38.

Arntzen JW, Espregueira Themudo G. 2008. Environmental parameters that determine species geographical range limits as a matter of time and space. Journal of Biogeography 35:1177-1186.

Attum O, Lee M, Roe JH, Kingsbury BA. 2008. Wetland complexes and upland-wetland linkages: landscape effects on the distribution of rare and common wetland reptiles. Journal of Zoology 275:245-251.

Austin MP 2002. Spatial prediction of species distribution: an interface between ecological theory and statistical modelling. Ecological Modelling 157:101-118

Austin MP, Van Niel KP. 2011. Improving species distribution models for climate change studies: variable selection and scale. Journal of Biogeography 38:1-8.

Baselga A, Lobo JM, Svenning JC, Araujo MB. 2012. Global pattern in the shape of species geographical ranges reveal range determinants. Journal of Biogeography 39:760-771.

Brotons L, Thuiller W, Araujo MB, Hirzel AH. 2004. Presence-absence versus presence-only modelling methods for predicting habitat suitability. Ecography 4:437-448.

Brown JH, Stevens GC, Kaufman DM. 1996. The geographic range: size, shape, boundaries, and internal structure. Annual Review of Ecology, Evolution, and Systematics 27:597-623

Chefaoui RM, Lobo JM. 2008. Assessing the effects of pseudo-absences on predictive distribution model performance. Ecological Modelling 210:478-486.

Compton BW, Beaudry F, McGarigal K, Sievert PR. 2007. Habitat Modeling for Blanding's Turtle (*Emydoidea blandingii*) in the Northeast. Final Report for the Northeast Blanding's Turtle Working Group.

Congdon JD, Graham TE, Herman TB, Lang JW, Pappas MJ, Brecke BJ. 2008*. Emydoidea blandingii* (Holbrook 1838)-Blanding's Turtle. In: Rhodin, A.G.J., P.C.H Pritchard, P.P. van Dijk, R.A. Saumure, K.A. Buhmann, and J.B. Iverson (Eds.) Conservation Biology of Freshwater Turtles and Tortoises: A Compilation Project of the IUCN/SSC Tortoise and Freshwater Turtle Specialist Group. Chelonian Research Monographs No.5, pp. 015.1-015.12

Crall AW, Jarnevich CS, Panke B, Young N, Renz M, Morisette J. 2013. Using habitat suitability models to target invasive plant species surveys. Ecological Applications 23:60-72.

DeCatanzaro R, Chow-Fraser P. 2010. Relationship of road density and marsh condition to turtle assemblage characteristics in the Laurential Great Lakes. Journal of Great Lakes Research 36:357-465.

Elith J, Graham CH, Anderson RP, Dudik M, Ferrier S, Guisan A, Hijmans RJ, Huettmann F, Leathwick JR, Lehmann A, Li J, Lohmann LG, Loiselle BA, Manion G, Moritz C, Nakamura M, Nakazawa Y, Overton JM, Peterson AT, Phillips SJ, Richardson K, Scachetti-Pereira R, Schapire RE, Soberon J, Williams S, Wisz MS, Zimmermann NE. 2006. Novel methods improve prediction of species' distributions from occurrence data. Ecography 29:129-151.

Elith J, Graham CH. 2009. Do they? How do they? Why do they differ? On finding reasons for differing performance of species distribution models. Ecography 32:66-77.

Elith J, Phillips SJ, Hastie T, Dudik M, Chee YE, Yates CJ. 2011. A statistical explanation of MaxEnt for ecologists. Diversity and Distributions 17:43-57.

Engler R, Guisan A, Rechsteiner L. 2005. An improved approach for predicting the distribution of rare and endangered species from occurrence and pseudo-absence data. Journal of Applied Ecology 41:263-274.

Ernst, C.H. and J.E. Lovich, 2009. Turtles of the United States and Canada. 2nd Edition Johns Hopkins University Press, Baltimore, Maryland.

Eskildsen A, le Roux PC, Heikkinen RK, Hoye TT, Kissling WD, Poyry J, Wisz MS, Luoto M. 2013. Testing species distribution models across space and time: high latitude butterflies and recent warming. Global Ecology and Biogeography 22:1293-1303.

ESRI (Environmental Systems Resource Institute). 2013. ArcMap 10.2.1. ESRI, Redlands, California.

Franklin J. 2009. Mapping Species Distributions; Spatial Inference and Prediction. Cambridge University Press: New York.

Gaston KJ. 2009. Geographic range limits: achieving synthesis. Proceedings of the Royal Society 276:1395-1406.

Gawler SC. 2008. Northeastern terrestrial wildlife habitat classification. A report to the Virginia Department of Game and Inland Fisheries on behalf of the Northeast Association of Fish and Wildlife Agencies and the National Fish and Wildlife Foundation.

Geber MA. 2008. To the edge: studies of species' range limits. New Phytologist 178:228-230.

Gesch D, Oimoe, M, Greenlee S, Nelson C, Steuck M, Tyler D. 2002. The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing 68:5-11.

Gesch DB. 2007. The National Elevation Dataset, in Maune D ed. Digital Elevation Model Technologies and Applications: The DEM User's Manual, 2nd Edition: Bethesda, Maryland, American Society for Photogrammetry and Remote Sensing: 99-118.

Gibbons, JW, Scott DE, Ryan TJ, Buhlmann KA, Tuberville TD, Metts BS, Greene JL, Mills T, Leiden Y, Poppy S, Winne CT. 2000. The global decline of reptiles, deja vu amphibians. Bioscience 50:653-66.

Gu W, Swihart RK. 2004. Absent or undetected? Effects of non-detection of species occurrence on wildlife-habitat models. Biological Conservation 116:195-203.

Guisan A, Zimmermann NE. 2000. Predictive habitat distribution models in ecology. Ecological Modelling 135:147-186.

Guisan A, Edwards TC, Hastie TJ. 2002. Generalized linear and generalized additive models in studies of species distributions: setting the scene. Ecological Modelling 157:89-100.

Heikkinen RK, Luoto M, Virkkala R, Pearson RG, Korber JH. 2011. Biotic interactions improve prediction of boreal bird distributions at macro-scales. Global Ecology and Biogeography 16:754-763.

Hijmans RJ, Cameron SE, Parra JL, Jones PG, Jarvis A. 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25:1965-1978.

Holt RD, Keitt TH. 2005. Species' borders: a unifying theme in ecology. Oikos 108:3-6.

Hortal J, Jimenez-Valverde A, Gomez JF, Lobo JM, Baselga A. 2008. Historical bias in biodiversity inventories affects the observed environmental niche of species. Oikos 117:847-858.

Jimenez-Valverde A, Lobo JM. 2007. Threshold criteria for conversion of probability of species presence to either-or presence-absence. Acta Oecologica 31:361-369.

Jimenez-Valverde A, Lobo JM, Hortal J. 2009. The effect of prevalence and its interaction with sample size on the reliability of species distribution models. Community Ecology 10:196-205.

Jin S, Yang L, Danielson P, Homer C, Fry J, Xian G. 2013. A comprehensive change detection method for updating the National Land Cover Database to circa 2011. Remote Sensing of Environment 132:59-175.

Johnson G, Crockett T. 2009. Distribution, population structure and habitat relationships of Blanding's turtle populations in northern New York. Final Report prepared for the Endangered Species Unit, New York State Department of Environmental Conservation.

Johnson G, Conrad L. 2012. Distribution of wetland-dependent amphibians and reptiles of greatest conservation need in northern New York- an application of the New York State Herps Atlas. State University of New York at Potsdam. Final report to the New York State Department of Environmental Conservation.

Kery M. 2002. Inferring the absence of a species: a case study of snakes. The Journal of Wildlife Management 66:330-338.

Khatchikian C, Sangermano F, Kendell D. 2011. Evaluation of species distribution model algorithms for fine-scale container-breeding mosquito risk prediction. Medical and Veterinary Entomology. 25:268-275.

Liu C, Berry PM, Dawson TP, Pearson RG. 2005. Selecting threshold of occurrence in the prediction of species distributions. Ecography 28:385-393.

MacKenzie DI, Nichols JD, Lachman GB, Droedge S, Royle JA, Langtimm CA. 2002. Estimating site occupancy rates when detection probabilities are less than one. Ecology 83:2248-2255

Manel S, Williams HC, Ormerod SJ. 2001. Evaluating presence-absence models in ecology: the need to account for prevalence. Journal of Applied Ecology 38:921-931.

Marchand MN and Litvaitis JA. 2004. Effects of habitat features and landscape composition on the population structure of a common aquatic turtle in a region undergoing rapid development. Conservation Biology 18:758-676.

Marini MA, Barbet-Massin M, Lopes LE, Jiguet F. 2010. Predicting the occurrence of rare Brazilian birds with species distribution models. Journal of Ornithology 151:857-866.

Mateo-Tomas P and Olea PP. Combining scales in habitat models to improve conservation planning in an endangered vulture. Acta Oecologica 35:489-498

McCluskey EM, Mockford SW, Sands K, Herman TB, Johnson G, Gonser RA. 2015. Population structure of Blanding's turtle (*Emydoidea blandingii*) in New York. Journal of Herpetology in press.

Merow C, Smith MJ, Silander Jr. JA. 2013. A practical guide to MaxEnt for modeling species' distributions: what it does, and why inputs and setting matter. Ecography 36:1058-1069.

Millar CS and Blouin-Demers G. 2011. Spatial ecology and seasonal activity of Blanding's Turtles (*Emydoidea blandingii*) in Ontario, Canada. Journal of Herpetology 45:370-378.

Millar CS and Blouin-Demers G. 2012. Habitat suitability modelling for species at risk is sensitive to algorithm and scale: a case study of Blanding's turtle, *Emydoidea blandingii*, in Ontario, Canada. Journal of Nature Conservation 20:18-29.

Millennium Ecosystem Assessment. 2005. Ecosystems and Human Well-being: Wetlands and Water Synthesis. World Resources Institute, Washington, DC.

Mockford SW, Herman TB, Snyder M, Wright JM. 2007. Conservation genetics of Blanding's turtle and its application in the identification of evolutionarily significant units. Conservation Genetics 8:209-219.

Olivero AP, Anderson MG. 2008. Northeast aquatic habitat classification system. The Nature Conservancy, Eastern Regional Office, Boston, MA.

Pearson RG, Dawson TP, Liu C. 2004. Modelling species distribution in Britain: a hierarchical integration of climate and land-cover data. Ecography 27:285-298.

Pearson RG. 2007. Species' distribution modeling for conservation educators and practitioners. Synthesis. American Museum of Natural History.

Pearson RG, Raxworthy CJ, Nakamura M, Townsend Peterson A. 2007. Predicting species distributions from small numbers of occurrence records: a test case using cryptic geckos in Madagascar. Journal of Biogeography 34:102-117.

Phillips SJ, Anderson RP, Schapire RE. 2006. Maximum entropy modeling of species geographic distributions. Ecological Modelling 190:231-259.

Phillips SJ, Dudik M. 2008. Modeling of species distributions with Maxent: new extensions and a comprehensive evaluation. Ecography 31:161-175.

Quesnelle PE, Fahrig L, Lindsay KE. 2013. Effects of habitat loss, habitat configuration and matrix composition on declining wetland species. Biological Conservation 160:200-208.

R Core Team (2014). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Randin CF, Dirnbock T, Dullinger S, Zimmermann NE, Zappa M, Guisan A. 2006. Are niche-based species distribution models transferable in space? Journal of Biogeography 33:1689-1703.

Refsnider JM and Linck MH. 2012. Habitat use and movement patterns of Blanding's turtles (*Emydoidea blandingii*) in Minnesota, USA: a landscape approach to species conservation. Herpetological Conservation and Biology 7:185-195.

Rizkalla CE, Swihart RK. 2006. Community structure and differential responses of aquatic turtles to agriculturally induced habitat fragmentation. Landscape ecology 21:1361-1375.

Ross, A.M., and G. Johnson. 2013. DRAFT Recovery Plan for New York State Populations of the Blanding's turtle (Emydoidea blandingii). New York State Department of Environmental Conservation, Albany, New York.

Rupprecht F, Oldeland J, Finckh M. 2011. Modelling potential distribution of the threatened tree species *Juniperus oxycedrus*: how to evaluate the predictions of different modelling approaches? Journal of Vegetation Science 22:647-659.

Seabrook L, McAlpine C, Rhodes J, Baxter G, Bradley A, Lunney D. 2014. Determining range edges: habitat quality, climate or climate extremes? Diversity and Distributions 20:95-106.

Segurado P, Araujo MB. 2004. An evaluation of methods for evaluation species distributions. Journal of Biogeography 31:1555-1568.

Sexton JP, McIntyre PJ, Angert AL, Rice KJ. 2009. Evolution and ecology of species range limits. Annual Review of Ecology, Evolution, and Systematics 40:415-436.

Swets JA. 1988. Measuring the accuracy of diagnostic systems. American Association for the Advancement of Science 240:1285-1293.

Tarkhnishvili D, Serbinova I, Gavashelishvili A. 2009. Modelling the range of Syrian spadefoot toad (*Pelobates syriacus*) with combination of GIS-based approaches. Amphibia-Reptilia 30:401-412.

Thuiller W, Richardson DM, Pyseks P, Midgley GF, Hughes GO, Rouget M. 2005. Niche-based modelling as a tool for predicting the risk of alien plant invasions at a global scale. Global Change Biology 11:2234-2250.

USDA National Agricultural Statistics Service Cropland Data Layer. 2010. Published crop-specific data layer [Online]. Available at http://nassgeodata.gmu.edu/CropScape/ (accessed July 16, 2014; verified July 16, 2014). USDA-NASS, Washington, D.C.

U.S Census Bureau. 2013 TIGER/Line Shapefiles [machine-readable data files]. Geography Division, (accessed July 16, 2014).

U. S. Fish and Wildlife Service. 1983. National Wetlands Inventory website. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. http://www.fws.gov/wetlands/

Wang WC, Lo NJ, Chang WI, Huang KY. 2012. Modeling spatial distributions of a rare and endangered plant species (*Brainea insignis*) in central Taiwan. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXIX-B7:241-246.

Werner, Jr., W.E. 1959. Amphibians and reptiles of the Thousand Islands region, New York. Copeia 1959:170-172.

Williams-Tripp M, D'Amico FJN, Page C, Bertrand A, Nemoz M, Brown JA. 2012. Modeling rare species distribution at the edge: the case of the vulnerable endemic Pyrenean desman in France. The Scientific World Journal 2012:1-6.

Wintle BA, Elith J, Potts JM. 2005. Fauna habitat modelling and mapping: A review and case study of the Lower Hunter Central Coast region of NSW. Austral Ecology 30:719-738.

Section 2

Species Distribution Modeling of Blanding's Turtle in the Massena Great Lakes Area of Concern

Habitat suitability predictions for the St. Lawrence River Valley

Our results support a priori observations that the range of the *E. blandingii* in northern New York State is limited by powerful environmental drivers. We found that elevation was the most important overall predictor of *E. blandingii* habitat suitability at both 250 m and 8,000 m scales (see Section 1). Statistical evaluation of model fit and validation in combination with visual assessment of predictive maps indicates that the presence-only machine learning method Maxent is better at predicting habitat suitability for *E. blandingii* both within the known range and outside it. For this reason only Maxent predictions will be discussed in this section.

Maxent identified a pronounced range edge for *E. blandingii* in the St. Lawrence River (SLR) Valley that is primarily associated with elevation (Figure 2.1). Habitat suitability predictions indicate that populations are concentrated near the SLR and suitable habitat is not uniform across the valley; the models identified more than half of it as unsuitable for *E. blandingii* (Figure 2.1). Maxent predictions corresponded closely with *E. blandingii* distribution records: most turtle records are in areas of predicted high habitat suitability and there are very few records in areas predicted as low suitability. *E. blandingii* in the SLR Valley seem to be occupying all suitable habitat. This also suggests that the sampling strategy in the SLR Valley has successfully identified all suitable *E. blandingii* habitat without the guidance of models. It is thus unlikely that additional, unknown populations exist in the valley. Efforts to conserve *E. blandingii* should be focused on known population locations.

Maxent identified a pronounced gap between *E. blandingii* populations in the Indian River Lakes ecoregion (between Jefferson and St. Lawrence Counties). This gap has been empirically observed by Dr. Glenn Johnson in his trapping efforts and also has been identified in recent genetic studies (McCluskey et al. 2015). Based on our findings, this gap is likely caused by temperature gradients, *E. blandingii* prefer cooler areas of the valley; thus habitat management to increase *E. blandingii* occupancy in this region is not recommended. McCluskey et al. however concluded that genetic diversity for the SLR Valley is relatively high and that overall the valley exhibited low levels of differentiation so gene flow between these “isolated” populations is still occurring. This finding has important local conservation implications in that attempting to establish new populations inside the gap may be impractical because of unsuitable habitat variables. However maintaining wetland and forest corridors can continue to support migration and genetic flow.

At the 8,000 m scale, Maxent also identified a gap between the Lisbon and Louisville populations. This gap was also personally observed by Dr. Glenn Johnson but not by McCluskey et al. (2015). The Maxent models identified higher stream density and lower coverage of shrub/scrub wetlands in the area of the gap as variables potentially responsible for this break. This finding suggests that habitat restoration or alteration, by increasing wetland coverage or connecting wetlands, may be a worthwhile strategy to link these two populations. Land acquisitions and conservation easements to conserve existing occupied habitat or potential habitat within this region would also help to conserve the existing populations and potentially expand their distributions into the currently unoccupied area.

Our results, while providing distribution information, say nothing about population dynamics in the SLR Valley. Additional demographic information and genetic studies would help clarify whether occupied patches are population sources or sinks, whether habitat patches are connected by genetic flow, and explain other micro-patterns (Brown et al. 1996; Gaston 2009).

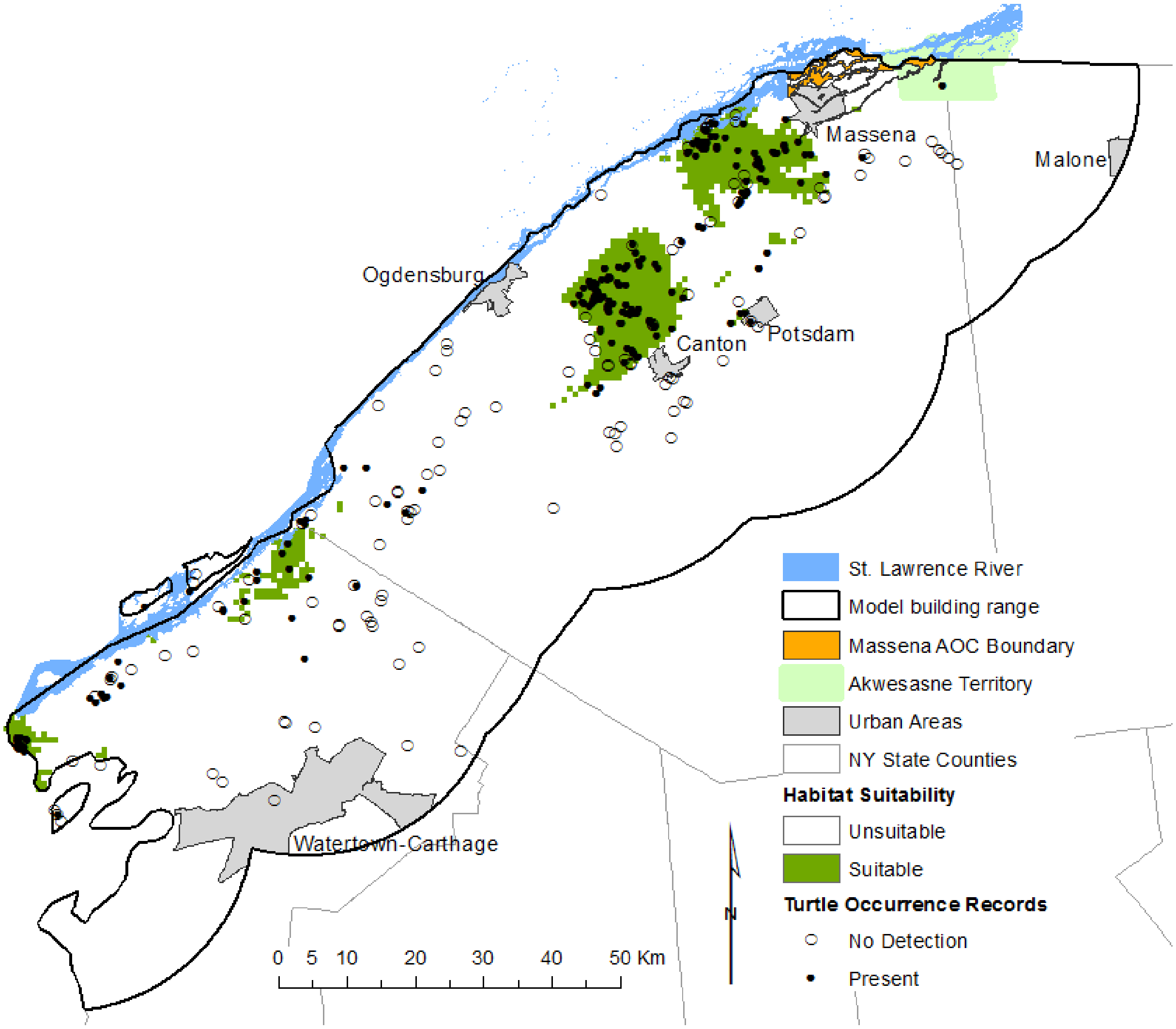


Figure 2.1: Probability of occurrence of *E. blandingii* across the St. Lawrence River valley using Maxent with an 8,000 m buffer. Green areas indicate high probability of occurrence. Survey locations are overlapped with model predictions.

In the area of the Massena AOC, the 250 m scale models predicted scattered patches of high suitability habitat, some falling within the city limits of Massena and within the AOC boundary (Figure 2.2A). At this scale low road density was the second most important predictor of high habitat suitability (see Section 1). Predicted areas were generally away from roads and also often overlapped with freshwater wetlands identified by the National Wetland Inventory (NWI; USFS 1983). At this fine scale, model predictions can help identify areas where habitat alteration and restoration may be desirable as part of Habitat Improvement Projects (HIPs) or other local conservation and management plans. The 8,000 m scale models identified a very large area of high habitat suitability in the region of Coles Creek and Wilson Hill Wildlife Management Area (WMA) (Figure 2.2B). This prediction corresponds to a high density of turtle occurrence records in this area. Models at this scale identified low stream density, high forested/shrub cover, and high hardwood cover as important variables to *E. blandingii* (see Section 1). Conservation, restoration, and management efforts should continue in this area and consider these variables when developing future projects. At neither scale did Maxent predict high suitability habitat within the Mohawk Territory of Akwesasne.

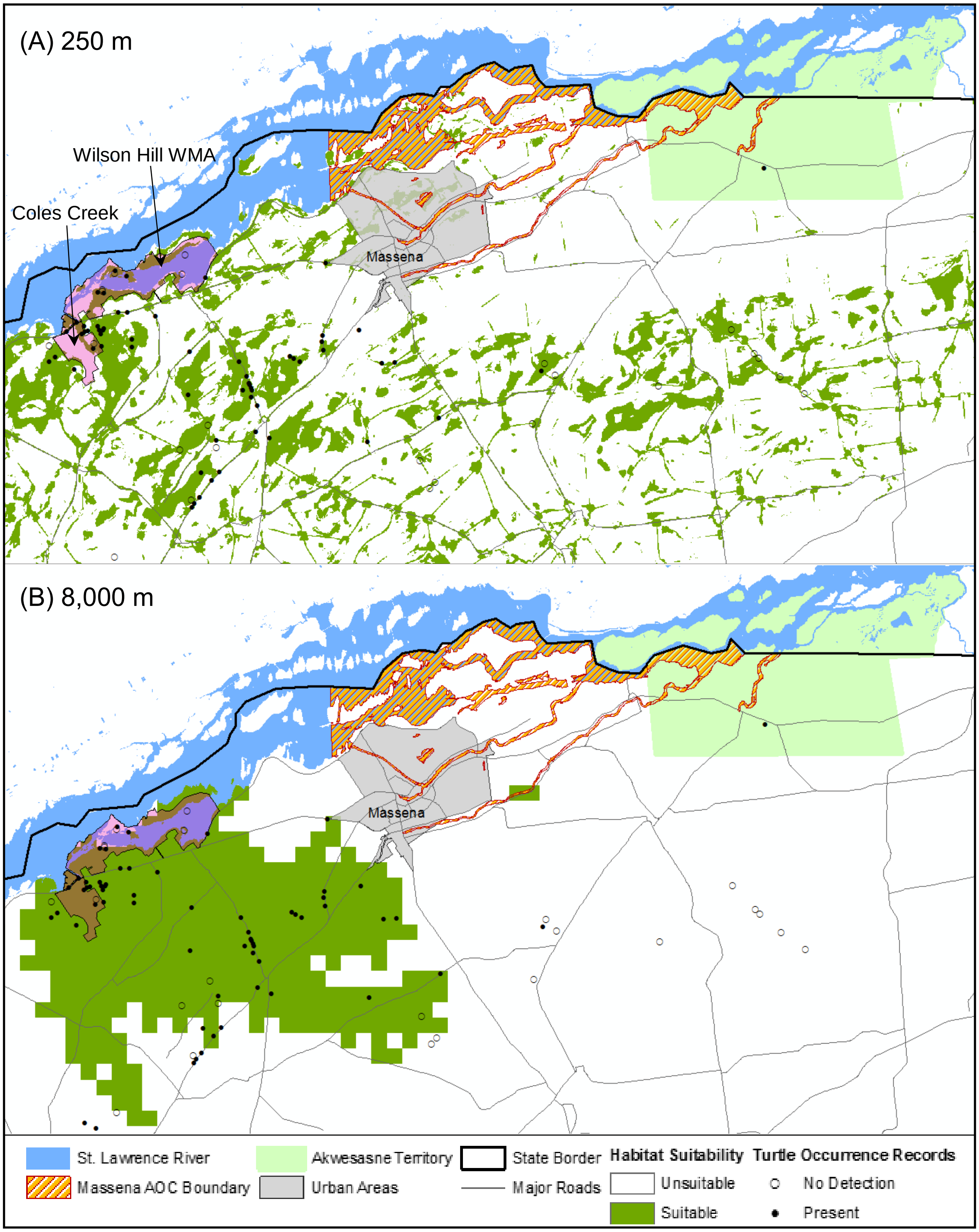


Figure 2.2: Probability of occurrence of *E. blandingii* near the Massena AOC using (A) Maxent with a 250 m buffer and (B) 8,000 m buffer. Green areas indicate high probability of occurrence. Survey locations are overlapped with model predictions.

Projection outside the St. Lawrence River Valley

We chose to project our SDMs to the rest of New York State to evaluate their performance in relation to several disjunct known populations of *E. blandingii* in the state, and to evaluate whether there may be suitable habitat elsewhere that could be of conservation value for this species. Maxent projections for the 250 m and 8,000 m scales were almost identical, thus only results from the 8,000 m analysis are illustrated and discussed.

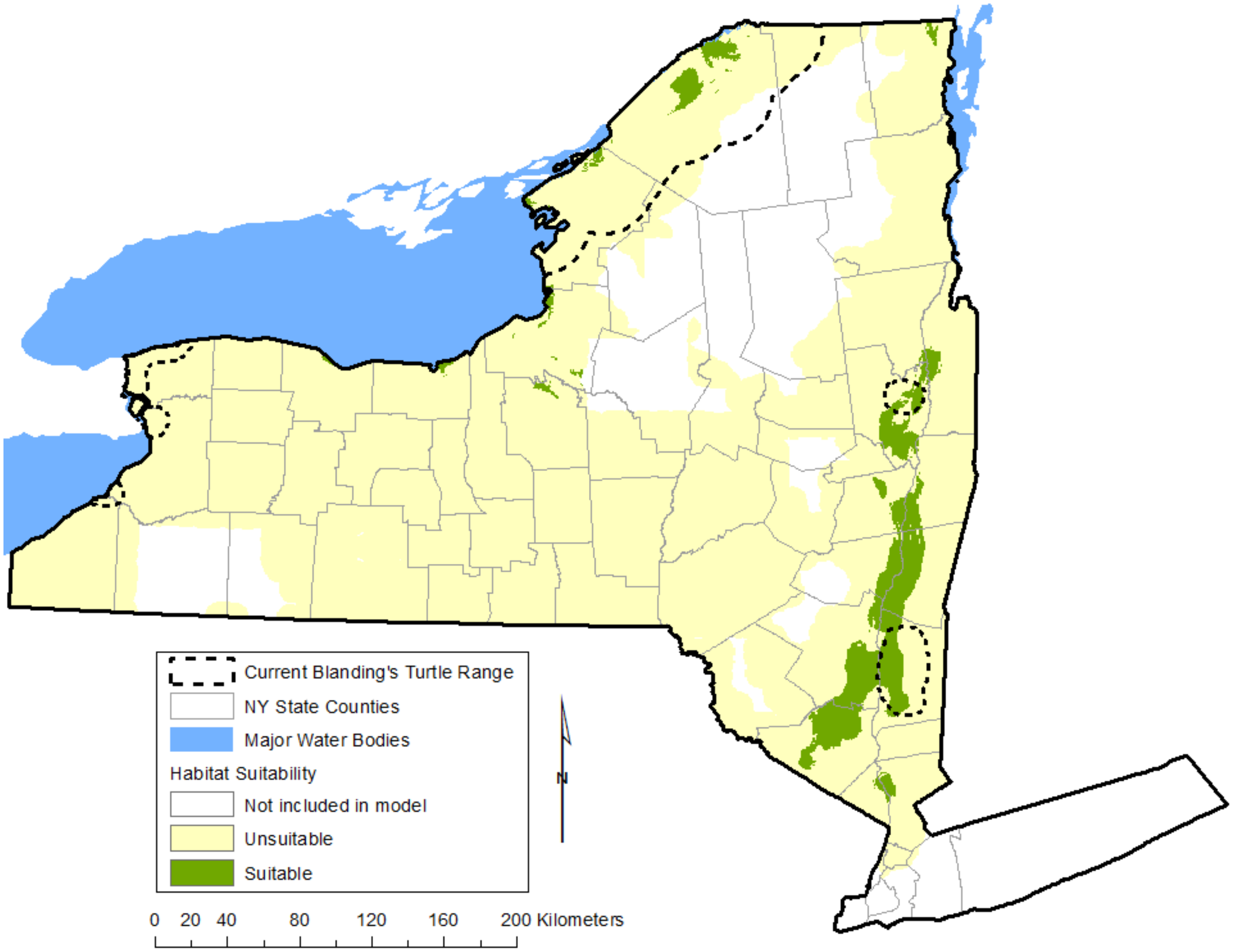


Figure 2.3: Projected areas of high habitat suitability for *E. blandingii* outside of the St. Lawrence River valley using Maxent with an 8,000 m buffer. The dotted line delineates the current known distribution of *E. blandingii* in New York State.

Maxent predicted suitable habitat for *E. blandingii* along the southeastern shore of Lake Ontario, on the western shore of Lake Champlain, and for a large section of the Hudson River valley, encompassing known disjunct populations in Saratoga and Dutchess Counties (Figure 2.3). Suitable but unoccupied habitat outside of the current range may indicate that (1) model projections are flawed, (2) undetected populations of turtles exist in these areas, (3) the species has not yet reached these suitable areas but may do so as a result of range expansion, or (4) populations have been extirpated from these areas due to historical habitat alteration, fragmentation, and degradation. We hypothesize that the regions of predicted extensive suitable habitat, the Hudson River and Lake Champlain Valleys and the Great Lakes lowlands, were once important components of the eastern *E. blandingii* range, but the historical extensive land cover and hydrological modification of the region for agriculture and industry resulted in region-wide extirpation, leaving as residuals the current small, disjunct populations. Many of our model building components were based on older GIS layers (circa 1980's) and likely do not represent the most current ecological structure, but rather a more historical perspective. McCluskey et al. (2015), observed that Dutchess and Saratoga populations are genetically isolated from the SLR Valley populations and each other. Even though the areas identified by our model may have once served as corridors for genetic flow, this no longer seems to be the case. It is thus unlikely that additional unknown populations of *E. blandingii* will be found in the areas projected by Maxent. The only exception to this may be the Lake Champlain Valley. This area was projected as suitable in every version of every modeling algorithm used in this study. This repeatability gives validation to these projections. We are unaware of any trapping efforts that have been performed in this area and therefore future trapping surveys may benefit from focusing there.

Conservation Implications

The Northeast Blanding's Turtle Working Group (NEBTWG) and the New York Department of Environmental Conservation (NYDEC) are in the process of developing conservation and recovery plans for *E. blandingii* in the northeast and in New York State. Various Habitat Improvement Projects (HIPs) funded by the FERC-relicensing agreement (2003) with the New York Power Authority (NYPA) and targeted towards *E. blandingii* are also in the process of being completed in the SLR Valley. The information provided by the species distribution models can be used to refine the steps taken to conserve, restore, and manage *E. blandingii* habitat, both locally and regionally. Future survey locations and protocols can also be targeted using these models.

The results of our models indicate that elevation is the dominant variable influencing the distribution of *E. blandingii* in the SLR Valley. This is an indication that the range edge of *E. blandingii* in northeastern New York is likely defined by a natural barrier rather than by habitat destruction. This finding does not however negate the fact that habitat destruction, nest predation, and road crossings are some of the severe threats to *E. blandingii* populations. What it does imply is that even when such threats are mitigated, conservation efforts to extend the range edge boundary via habitat management or population translocations are likely to be ineffective because of the limits imposed by the natural topographic barrier. One next step we can recommend is to fully evaluate the apparent gaps in *E. blandingii* occurrence with Jefferson and St. Lawrence County border area of the Indian River Lakes Region to verify whether this gap does exist, and if so whether it is natural or due to human activities that could potentially be mitigated.

Habitat suitability predictions in the SLR Valley indicate that some populations may be undergoing fragmentation possibly due to a combination of factors, including some not included in our models such as competition or limited dispersal ability. Our model results indicate that habitat characteristics, such as availability of forested/shrub wetlands and stream density may be responsible for the isolation of some populations. These variables are likely associated with habitat degradation and can be targeted as part of conservation plans. The predicted distribution patters produced by our models can be augmented by additional demographic or genetic studies. There are two obvious next steps for evaluating the extent to which habitat fragmentation and other factors limit *E. blandingii* in the SLR Valley. These include (1) a careful assessment of whether and how road barriers and road mortality limit populations, and identification and implementation of potential mitigation measures for any detected negative effects of roads on the species. (2) A study of whether availability of suitable nesting habitat limits the distribution of *E. blandingii* in the SLR Valley. If so, habitat management measures such as created nesting habitat or habitat augmentation and predator control at existing nesting habitat may be worthwhile. Our SDM model can indicate the most promising areas to focus these investigations.

Because SDM projections are inherently uncertain, one must be cautious making conservation decisions based on their forecasts. Nevertheless, our results indicate that projections can provide clues to the historical species distribution and potential for species range expansion. Areas identified as suitable by projections can be targeted for future surveys and even evaluated as candidates for habitat management and population translocations to connect disjunct populations.

References

Brown JH, Stevens GC, Kaufman DM. 1996. The geographic range: size, shape, boundaries, and internal structure. Annual Review of Ecology, Evolution, and Systematics 27:597-623

Gaston KJ. 2009. Geographic range limits: achieving synthesis. Proceedings of the Royal Society 276:1395-1406.

McCluskey EM, Mockford SW, Sands K, Herman TB, Johnson G, Gonser RA. 2015. Population structure of Blanding's turtle (*Emydoidea blandingii*) in New York. Journal of Herpetology in press.

U. S. Fish and Wildlife Service. 1983. National Wetlands Inventory website. U.S. Department of the Interior, Fish and Wildlife Service, Washington, D.C. http://www.fws.gov/wetlands/